REVIEW

Coevolution of host-parasite associations and methods for studying their cophylogeny

A Filipiak[a], K Zając[b], D Kübler[b], P Kramarz[b]

[a]Institute of Plant Protection - National Research Institute, Władysława Węgorka 20, 60-318 Poznań, Poland
[b]Institute of Environmental Sciences, Jagiellonian University, Gronostajowa 7, 30-387 Kraków, Poland

Abstract

Coevolution may be defined as the process of reciprocal, adaptive genetic change in two or more species. Host-parasite interactions play an important role in the evolutionary ecology. The host phylogeny is independent, and the phylogeny of the parasite depends to some extent on the host. This review provides a description of several different methods for studying host-parasite relationships, along with a description of the underlying models and theoretical background for each. It also shows the possible applications of different methods and describes the advantages and drawbacks of different techniques.

Key Words: coevolution; host-parasite interactions; cophylogenetic methods; event-based methods; global fit methods

Introduction

One of the most fascinating topics in evolutionary biology is the coevolutionary race between hosts and parasites, driven by antagonistic interactions which lead to reciprocal adaptations. Most parasites have shorter generation times and are more numerous than their hosts, and therefore evolve faster, adapting to and increasing the selective pressure on the host. When parasites adapt to the most common host genotype, rare host genotypes become favored by selection, resulting in changes in genotype frequencies in host populations. This negative frequency dependent selection results in fluctuations in the relative frequencies of genotypes and also maintains genetic diversity. For example, Decaestecker *et al.* (2007) studied the infectivity of the parasitic bacterium *Pasteuria ramosa* to its host *Daphnia magna* using resting stages of hosts and parasites from different layers of pond sediments. By exposing hosts to parasites from different layers, the researchers showed that bacteria from a given sediment layer were most easily able to infect hosts that originated from the same layer, indicating that the parasites were adapted to the most-frequently encountered host genotypes. Knowledge about the influences of hosts on parasites, and vice versa,

becomes particularly important because of its relevance and applicability to many current problems in biology, such as controlling the spread of diseases, predicting the influence of expanding invasive species, or identifying and developing biological control agents (de Vienne *et al.*, 2013).

The host-parasite association is also a representative model for studying adaptation. Reconstruction of a parasite's ancestry through the study of its ancestral host is far less complex than reconstructing the entire ecosystem for free-living species (Paterson and Banks, 2001). In interpreting relationships between coevolving lineages, however, it is important to adopt a phylogenetic perspective. Because of this, in this kind of study it is necessary to reconstruct the patterns of relatedness between related hosts and their respective parasites. When we think about associations between hosts and parasites, we can find an analogy in vicariance biogeography. In this metaphor, the host can be considered an 'area' for the parasite to exploit, with host speciation as a 'vicariance event' (Page and Charleston, 1998). The relationship between host and parasite can also be seen as analogous to that between gene and species trees (Page and Charleston, 1998). In the past, parasites were more or less considered equivalent to host phenotypic characters and thus, the parasite's phylogeny was believed to mirror that of the host (Desdevises, 2007), a pattern known as Fahrenholz's rule (Brooks 1985; Huelsenbeck *et al.*, 2000). This rule, formulated in 1913, helped to develop knowledge in the area of host-parasite

Corresponding author:
Paulina Kramarz
Institute of Environmental Sciences
Jagiellonian University
Gronostajowa 7, 30-387 Kraków, Poland
E-mail: paulina.kramarz@uj.edu.pl

coevolution; however, further studies have shown that most host and parasite phylogenies are partly incongruent (Brooks and McLennan, 1993; Paterson and Banks, 2001; Desdevises *et al*., 2002). Because, in general, a parasite's phylogeny only imperfectly mirrors that of the host, Paterson and Banks (2001) proposed to reformulate Farenholz's rule. Even when we take into account phylogenetic relatedness and phenotypic or ecological similarity, however, comparative analyses may still lead to confounding results because of limitations in the ability of current methodologies to model all possible evolutionary scenarios.

During the common history of hosts and parasites, five basic categories of events can shape the coevolutionary process. These are: cospeciation, host switching, duplication of parasite lineages, sorting events, and inertia (Page and Charleston, 1998; Paterson and Banks, 2001; Desdevises, 2007). Cospeciation events occur when host and parasite species diverge together (*e.g.*, during host isolation). In the majority of cases, it takes place in two-species systems, but there is some evidence for cospeciation involving three species, specifically, in a symbiotic system made up of a fly, a nematode, and a plant (Nelson *et al*., 2014). Host-switching occurs when a parasite establishes on a new host and diverges from the original form as selection favours adaptations to the new host (Imber, 1985; Brooks and McLennan, 1993; Shaw, 1994; Thompson, 1994; Hoberg *et al*., 1997; Paterson and Gray, 1997; Norton and De Lange, 1999; Ricklefs and Fallon, 2002; Susoy and Herrmann, 2014). Duplication takes place when only a parasite lineage, and not its host, speciates. Sorting events may include extinction or 'missing the boat', in which speciation as the result of a founder effect leads to the loss of a parasite in a particular lineage of host (Johnson *et al*., 2003). Finally, 'inertia' refers to the reverse of duplication, when a parasite species remains the same despite host speciation.

In order to take into account different evolutionary events when comparing host and parasite trees, researchers must make use of complex analytical and numerical methods. For example, many studies have attempted to identify instances of coevolution, including cospeciation between hosts and parasites, using phylogenies based on DNA sequences (Dowton and Austin, 1995; Huelsenbeck *et al*., 1997; Huelsenbeck *et al*., 2000; Dixon, 2002). Relatively, several methods were developed to reconstruct hypothetical coevolutionary scenarios and to assess levels of cospeciation. The development of new methods along with advances in computing has brought an improvement in the power and precision of coevolutionary studies (Charleston, 1998; Page and Charleston, 1998; Cornell *et al*., 1999; Paterson and Banks, 2001; Meier-Kolthoff *et al*., 2007). This review provides a description of several different methods for studying host-parasite relationships, along with a description of the underlying models and theoretical background for each. It also shows the possible applications of different methods and describes the advantages and drawbacks of different techniques.

Theoretical background

Coevolution may be defined as the process of reciprocal, adaptive genetic change in two or more species (Brooks, 1987; Woolhouse *et al*., 2002; Banks and Paterson, 2005), and the study of this phenomenon relies heavily on the examination of phylogenies of eco-connected groups. In the case of hosts and parasites, the host phylogeny is independent, and the phylogeny of the parasite depends to some extent on the host (Stevens, 2004). This system is unique among coevolving associations because of its intimate nature and the strong selective pressures that each population can exert on the other. The process of host-parasite coevolution plays an important role in maintaining genetic variation, which may lead to changes in patterns of biological diversity (Thompson, 1994; Gandon *et al*., 2002; Woolhouse *et al*., 2002; Tellier and Brown, 2007; Forde *et al*., 2008). It is important to note, however, that coevolution should not be confused with the concepts of cospeciation, codivergence, or cophylogeny, which differ from coevolution in several important aspects.

Cospeciation involves the joint speciation of two or more species that are ecologically associated (*e.g.*, host-parasites). However, the speciation of symbionts may occur independently of host speciation, often through host shifts as the symbiont comes to occupy a new host environment in isolation from the ancestral lineage. Cospeciation is expected to have congruent phylogenies but also to have similar divergence times. Similar congruent topologies as seen in cospeciation could arise as a result of host switches followed by cospeciation events (or pseudocospeciation) but not have similar divergence times. Some authors caution against the use of 'coevolution' as a synonym for cospeciation because of the implication that short-term dynamics contributes directly to cospeciation in the long term, although the rationale underlying this idea and its potential implications have never been fully articulated (Page, 2003; Smith *et al*., 2008; de Vienne *et al*., 2013; Herrera *et al*., 2016).

Codivergence is the parallel divergence of ecologically associated lineages within two distinct phylogenies, and is one predicted outcome of coevolution. When sustained codivergence is accompanied by multiple examples of parallel phenotypic change, reciprocal coevolution can be considered a more probable mechanism than, for example, secondary, one-sided evolutionary change by one species to match its co-mimic, as follows. Codivergence may not prove coevolution in the strict sense. However, codivergence can be considered some of the strongest available evidence for coevolution (Joron and Mallet, 1998; Page, 2003; Hoyal Cuthill and Charleston, 2012).

Cophylogeny focuses on species associations (organisms tracking organisms, such as parasites and hosts or pollinators and flowering plants), molecular systematics (organisms or genes tracking genes) and historical biogeography (organisms tracking areas). Studies on cophylogeny stem from the observation that the diversification patterns over evolutionary time of tightly associated organisms, such as parasites and their hosts, are seldom

independent. Therefore, some degree of topological similarity, often termed congruence, between the phylogenies of the associated taxa is expected to occur. Congruence quantifies the extent to which each node in a given tree maps to a corresponding position in the other tree and perfect congruence can be interpreted as evidence for cospeciation, which may or may not result from coevolutionary mechanisms (Ronquist, 1997; Jousselin *et al.*, 2008; Baum and Johnson, 2010; Balbuena *et al.*, 2013).

Population genetics, natural selection in gene pools, adaptation, and isolating mechanisms have traditionally been considered by many biologists as significant factors in the evolutionary process. Hosts and their symbionts are involved in intimate physiological and ecological interactions. The impact of these interactions on the evolution of each partner depends on the time-scale considered. Coevolution might occur either in beneficiary relationships as mutualisms or symbioses, or in antagonistic relationships as host-parasite systems. Symbiotic associations are widespread phenomenon and these have had a significant impact on the ecology and evolution of many prokaryotes and eukaryotes. For hosts that are obligately dependent on the symbiont for survival, acquisition of the symbiont is a critically important life-history event for all members of each new host generation. To date, the coevolution of hosts and microbial symbionts from various environments has been analysed to disclose their symbiotic history (Paracer and Ahmadjian, 2000; Noda *et al.*, 2007; Dunlop *et al.*, 2012; Erler *et al.*, 2012; de Vienne *et al.*, 2013). Starr has developed a classification which was called the symbiotic continuum (Starr, 1975). This classification of symbiotic categories is based on the potential fitness, or reproductive ability, of the symbionts. Competition and mutualism are at opposite ends of the continuum, and neutralism is in the center. Competition between interacting species produces detrimental outcomes to both species, whereas mutualistic relationships increase the potential fitness of the symbionts. Thus, symbiosis means that two interacting organisms can influence each other's rates of survival and reproduction. Other continua in Starr's classification deal with the duration of the symbiosis, from transient to permanent; the physical contact between the symbionts, from incidental to close; and nutrition, from saprobic to biotrophic (Paracer and Ahmadjian, 2000). The popularly known symbiosis is lichen, which is a merger of algae or cyanobacteria and fungi. The cyanobacteria or algae provide the photosynthetic metabolism while the fungus can reorganize its membranes to sustain the lichen in extreme weather changes such as frozen tundra and desert rocks (Hird, 2010).

From an evolutionary perspective, host-parasite relationships may begin in conflict but eventually they move toward compromise, and the immune system plays a central role in the complex system of checks and balances. The basic protective strategy of an innate immune system is for the organism to constitutively produce generic receptors that recognize conserved patterns on different classes of pathogens to trigger an inflammatory response that limits pathogen invasion (Read, 1994; Cooper and Alder, 2006). Distinctions between self and non-self exist in all animals and show phylogenetic complexity and adaptive immune responses. Mounting an immune response is metabolically expensive; a host may compromise between the available resources for its growth and development and for defence (Behnke *et al.*, 1992). A balance between the beneficial and potential harmful effects of immune responses to infection has to be considered in terms of a series of evolutionary trade-offs. For the host it involves resistance, pathology, and loss of resources, and for the parasite it involves reproduction, immunogenicity, and pathogenicity. Compromise strategies have led to stable equilibria between many parasites and their hosts. Genetic structure, evolution of microparasites, mechanisms of pathogenesis, and the evolution of immune response have consequences for public health, treatment, and prevention. The acquired immune response results in the recovery of the host from a disease and is followed by the host acquiring a specific memory, with which it responds vigorously to an infection by the same parasite (Read, 1994; Paracer and Ahmadjian, 2000). Parasitologists have often described the phenomenon of premunition, which is the immunity of a host to reinfection following recovery from disease. The exact mechanism of premunition is not understood. Some parasites, however, have evolved novel strategies to counter a host's immune system and are able to establish long-term chronic infections (Paracer and Ahmadjian, 2000; Cooper and Adler, 2006).

There are many methods for the study of the phylogeny of host-parasite associations, so called cophylogenetic analyses. Generally, they are divided into two categories: "event-based methods" and "global fit methods". Event-based methods apply the five coevolutionary scenarios described above (*i.e.*, cospeciation, host switching, duplication of parasite lineages, sorting events, and inertia) to map the phylogenies of host and parasite. These methods are aimed at finding the most probable coevolutionary history of the associated taxa. Moreover, event-based methods have strong appeal because they promise to deliver the coevolutionary history of the associated taxa. The process model and the cost assignments of the method reveal its properties. Thus, it is straightforward to compare event-based methods and predict how they will perform when applied to particular problems. Three challenges are important in their application. First, well resolved phylogenies are required to obtain reliable results and even with a small number of taxa the number of equally parsimonious solutions can be exceedingly high. Then, event-cost methods are strongly dependent on a good estimation of the set of costs considered. And finally, given that not all the topological congruence between trees is necessarily a result of cospeciation, the precise reconstruction of coevolutionary history often requires additional data, such as the ages of the nodes, assumptions on the probability of the different events, consideration to the geological history of the areas involved and experimental evidence, such as reciprocal transplant experiments

(Hypsa, 2006; de Vienne *et al.*, 2007; Balbuena *et al.*, 2013). In this category we find Brooks' Parsimony Analysis, reconciliation analysis, cost-based methods, and probabilistic methods (Cruaud *et al.*, 2012; Balbuena *et al.*, 2013). Global fit methods instead of focusing on individual coevolutionary scenarios assess the congruence between host and parasite trees and also specify particular host-parasite associations that can help in deciphering the cophylogenetic structure (Desdevises, 2007). To some extent, the approach taken by global-fit methods is similar to statistical tests for congruence between two given trees. There is a clear need for this kind of methods because they afford large-scale cophylogenetic analyses for which the application of event-based counterparts becomes computationally prohibitive (Balbuena *et al.*, 2013). Distance-based methods provide a good example of this approach. The advantages and disadvantages of each method are presented below.

Methods for studying cophylogeny

1. Brooks' Parsimony Analysis (BPA; Brooks, 1981)

One of the first numerical methods aimed at examining cospeciation events was Brooks' Parsimony Analysis, developed in the early '80s. This method was proposed by Wiley on the basis of ideas developed by Brooks (Morrone and Crisci, 1995), and is based on the approach of mapping a parasite species and its phylogeny as characters and a transformation series, respectively, on the host phylogeny (Brooks, 1988). The protocol of the analysis includes identifying congruent and incongruent parts of the parasite and host phylogenies. The second step is mapping the parasite phylogeny onto the host phylogeny and estimating the fit measures. Among the many different phylogenies generated, the most parsimonious one is retained. In many cases, though, additional information is required to address incongruence in the host and parasite trees. That is the weakness of this analysis: although the number of overall ad hoc hypotheses is minimized, a large number of post hoc hypotheses is required. A suitable program for this method is Hennig86; the BPA method, once widely used, has now been displaced by other methods. In historical biogeography and coevolutionary analysis, it became evident that homology and homoplasy, as revealed by BPA and component analysis, had no straightforward interpretation in terms of evolutionary events such as dispersals and host shifts (Page, 1994; Charleston, 1998).

2. Reconciliation analysis (Page, 1990; 1994; Charleston, 1998)

Reconciliation methods use the same null hypotheses as BPA; however, these two methods differ in their essential approaches. Reconciliation methods are based on conditions determined *a priori* by the researcher. The parasite phylogeny is mapped onto the host phylogeny, and the best scenario is chosen using the criteria of the minimum number of events inferred or the least cost (de Vienne *et al.*, 2013).

The first software developed for use in reconciliation was Component. It estimates the best reconstruction scenario by minimizing the number of extinctions and intrahost speciation events and maximizing the number of cospeciation events. However, it completely ignores host switching (Page, 1994). The component performs several methods of comparison and can generate consensus trees, calculate the similarity between pairs of trees, and map one tree onto another. It has the ability to perform measurements such as partition metrics and quartet measures, which allow for the quantification of phylogenies between host and parasite. However, the biggest drawback of Component, as mentioned above, is that it does not include host switching as a coevolutionary event (Dowling *et al.*, 2003).

Instead, TreeMap (Page, 1994) includes all host switching scenarios, but does not model them correctly (Charleston, 1998). This method tries to reconcile host and parasite phylogenies by maximizing the number of cospeciations and minimizing the number of host-shift speciations. There are no constraints on the numbers of intrahost speciations or extinctions or on numbers of parasites present on internal nodes, so the number of parasites infecting ancestral host species or number of intrahost speciations may be assumed to be unreasonably high (Refrégier *et al.*, 2008). The complexity of the problem of multiple host switches and temporal incongruence in the coevolution of hosts and parasites inspired Charleston (1998) to provide a new algorithm, called Jungle. This method was implemented in the program TreeMap2, and it takes into consideration each hypothesized past association individually in order to find globally optimal solutions. The algorithm considers the costs of different phylogenetic events that may affect a parasite-host association, including multiple host switches, and gives more realistic coevolutionary scenarios than TreeMap does.

Another useful tool for reconciliation analysis is Tarzan, which allows the user to define the timing of the nodes in a parasite phylogeny and enables rapid analysis. Unfortunately, it has some disadvantages: for example, the solution proposed by Tarzan may not necessarily be optimal and sometimes it is not possible to find a solution even if it exists (Merkle and Middendorf, 2005).

The program Jane is slower than Tarzan, but has many advantages. It not only allows the user to define the timing of the nodes in a parasite phylogeny (like Tarzan does), but in a host phylogeny as well. In addition, the program enables the user to define different host-switch costs independently and the maximum permitted host-switch distance (Conow *et al.*, 2010).

3. Generalized parsimony or cost-based method (Ronquist, 1995)

Ronquist (1995) proposed a method of reconstructing host-parasite associations based on costs or weights related to the likelihood of occurrence of different kinds of coevolutionary events. The method involves the conversion of a host phylogeny into a cost matrix and allows for different evolutionary events, including host

switches. After that transformation, the ancestral states (hosts) on the parasite tree are optimized. The weight of switching events is specified relative to tracking events. Generalized parsimony algorithms are used to find the least-cost historical reconstruction of the host-parasite association. This method can be implemented in TreeFitter. It estimates the number of events of each type that could explain the observed congruence between the two phylogenies. Then, TreeFitter associates each event with the probability that it arose by chance, calculated by permutations of the host and/or parasite leaves on the phylogeny (de Vienne *et al.*, 2013). Its main advantage is that the costs of each event are set by the user (Ronquist, 1995). The main disadvantages are that cospeciation is considered to be the most parsimonious hypothesis, and there is no possibility to set cospeciation as costlier than host-shift speciation. Also, temporary non-compliance can appear, leading to erroneous conclusions.

4. Probabilistic methods (Huelsenbeck et al., 2000)

To test whether host and parasite phylogenies are identical Huelsenbeck *et al*. (2000) proposed two different null hypothesis tests. The first is based on maximum likelihood and apply a likelihood ratio test which evaluates the likelihood that the two phylogenies are identical versus the likelihood that they are not. The second is based on maximum posterior probability, uses Bayesian inference to directly calculate the posterior probabilities of the host and parasite phylogenies.

The maximum likelihood method uses standard statistical techniques for inferring probability distributions in order to assign probabilities to particular possible phylogenetic trees. It requires a substitution model to assess the probability of particular mutations; roughly, a tree that requires more mutations at interior nodes to explain the observed phylogeny will be assessed as having a lower probability. This is broadly similar to the maximum-parsimony method, but maximum likelihood allows additional statistical flexibility by permitting varying rates of evolution across both lineages and sites. In fact, the method requires that evolution at different sites and along different lineages be statistically independent (Simmons, 2012).

The Bayesian method is based on a simple stochastic model of host switching by a parasite in which there is an assumption that host switching events occur at a constant rate. The posterior probability density of the parameters of the model is evaluated numerically using a Markov chain Monte Carlo approach. The method is reliable, as it treats both host and parasite phylogenies as random variables and accounts for phylogenetic non-independence. However, its main weakness is that it does not include some important biological processes - it only considers host shifts and cospeciation. This analysis can be carried out using MrBayes (Ronquist and Huelsenbeck, 2003).

Although methods based on maximum likelihood and Bayesian inference have been specifically designed to study the cophylogeny of host and parasites, the applicability of these methods to cophylogenetic studies is limited because they are primarily intended for one-to-one associations, something that rarely occurs in nature (Balbuena *et al.*, 2013).

5. Distance-based methods

Another group of methods is based on statistical tests for congruence between host and parasite phylogenies. A comparison is made between the assumption that the two trees are independent (null hypothesis) and the probability of obtaining a certain level of congruence between the trees. In this kind of analysis there is the possibility of obtaining more reliable *a posteriori* interpretations. The large number of random trees that must be generated *de novo* for each new comparison of trees which is one of the weaknesses of these methods. To overcome this problem has been proposed a test of tree independence based on comparisons of the topological or genetic distances of the focal host-parasite association with a distribution of distances computed from a large number of randomly generated trees. Tests of independence have also been used to evaluate temporal congruence in the speciation histories of hosts and parasites (Kupczok and von Haeseler, 2009; de Vienne *et al.*, 2013). The distance-based method described by Hafner and colleagues (1994) uses alignments for specific loci from a host and a parasite to calculate distance matrices and then to test whether the host and the parasite have accumulated a similar number of genetic differences. After that analysis, a Mantel test is conducted to assess the significance of the correlation between the two matrices. This test can verify statistical non-independence, but it does not handle testing for phylogenetic non-independence (Hafner *et al.*, 1994).

In ParaFit (Legendre *et al.*, 2002), another distance-based method, either raw data or trees for a host and parasite can be converted into distance matrices. The ParaFit method adapts well to any kind of host-parasite association, including cases in which multiple parasites are associated with one host. It can then assess the contribution of each individual host-parasite association to the total congruence statistics (de Vienne *et al.*, 2013). However, the main drawback of this method is that it does not account for phylogenetic non-independence (Felsenstein, 1985).

A different kind of permutation test was proposed by Hommola and co-workers (2009). Just as in the ParaFit method, this method tests the null hypothesis that a host and parasite have evolved independently of each other using their phylogenetic trees and host-parasite association links. Both methods use permutation tests to determine whether cospeciation occurred. Moreover, Hommola's permutation test is based on the calculation of Pearson's correlation coefficients between host distances and parasite distances, considering all pairs of interacting hosts and parasites. (Hommola *et al.*, 2009; de Vienne *et al.*, 2013). Unlike ParaFit, this method does not evaluate the contribution of individual host-parasite links to the global cophylogenetic structure. In addition, this method differs from ParaFit in the

randomization procedure to test the significance of the global-fit statistic (Balbuena *et al*., 2013).

Schardl and co-workers (2008) proposed the MRCAlink (MRCA for Most Recent Common Ancestor) algorithm. This method identifies phylogenetically independent pairs between host and parasite trees and the reduced host and parasite matrices can then be compared. The method is less dependent on tree topologies, which often can be misleading, compared to tree reconciliation, and it crucially improves on phylogeny-independent methods such as ParaFit or the Mantel test by eliminating an extreme (but previously unrecognized) distortion of node-pair sampling (Schardl *et al*., 2008; de Vienne *et al*., 2013).

Recently described method, PACo (Procrustes Approach to Cophylogeny) is based on Procrustes analysis. This analysis is an extremely flexible technique used for displaying two or more multivariate datasets in their optimal superimposition. PACo provides a superimposition plot enabling a graphical comparison of the fit of the host-parasite associations. This test can be carried out with any pair of distance or dissimilarity matrices, *i.e.,* fully resolved host and parasite phylogenies are not required, and allows for multiple host-parasite associations and different number of hosts and parasites. PACo is also similar to ParaFit in that it uses the same three data matrices as input and converts the phylogenies to principal coordinates (PCo), and it is possible to assess the contribution of individual host-parasite associations to the global topological congruence. An important conceptual difference with the previous tests is that both ParaFit and test was proposed by Hommola and co-workers (2009) compare the host and parasite distance matrices and test for random association between the host and parasite taxa, whereas PACo explicitly tests the dependence of the parasite phylogeny upon the host phylogeny, because in the Procrustean superimposition, the parasite matrix is rotated and scaled to fit the host matrix (Balbuena *et al*., 2013).

As shown above, different methods of studying coevolution can vary fundamentally from each other or only in details. It is possible to distinguish two main categories of methods according to the null hypothesis that they test: event-based methods, in which coevolution is considered the most parsimonious explanation for congruence between host and parasite trees, and distance-based methods, which test the independence or similarity between trees or alignments (de Vienne *et al*., 2013).

A very useful categorization was proposed by Dowling *et al*. (2003). They distinguished two main classes of methods for studying host-parasite associations, *a priori* and *a posteriori* methods. Their deliberations were inspired by Van Veller and colleagues' (2002) suggestion that *a priori* and *a posteriori* methods for studying biogeographical patterns correspond to different research programs established on different ontological bases. By analogy, *a priori* and *a posteriori* methods for studying host-parasite associations, although they test the same null hypothesis, are based on different

ontologies. *A priori* methods (*e.g.*, reconciliation methods) modify the data to fit the null hypothesis of coevolution (parsimony is considered an *a priori* criterion) and therefore cannot be falsified. Instead, *a posteriori* methods (*e.g.*, BPA) provide the opportunity to falsify the null hypothesis when the data do not support it. In some cases *a priori* and *a posteriori* methods yield the same results and interpretations, but not always. If coevolution were influenced mostly by coevolution, an *a priori* method could be preferred. However, when we do not know if coevolution is the dominant event, *a posteriori* methods are superior; they avoid oversimplification of the results and false conclusions (Dowling *et al*., 2003). Moreover, de Vienne *et al*. (2013) showed that convincing cases of coevolution are rare and that cophylogenetic methods often overestimate their occurrence. Furthermore, based on a literature review and theoretical considerations, they concluded that parasite speciation by host shift is more common than coevolution (de Vienne *et al*., 2013). Thus, it is very important that cophylogenetic methods take host shifts and other evolutionary events into consideration. In addition, Fahrenholz's rule must be redefined as speciation via host shifts is at least as likely as coevolution (de Vienne *et al*., 2013). Moreover, incongruence between host and parasite phylogenies can occur despite a very intense cophylogenetic history. Similarly, congruence is not evidence of coevolution - congruence means little without additional supporting data (Paterson and Banks, 2001). It would be ideal if a testable method of reconstructing and interpreting all phylogenetic events existed; however, it is not available yet, and in studying host-parasite associations one has to make compromises.

Dowling and colleagues (2003) insisted that *a priori* methods such as reconciliation methods fail to satisfy three fundamental principles of phylogenetic systematics: the principle of total evidence (using all available evidence), the principle of maximum explanatory power (selecting the best *a posteriori* working hypothesis that best fits all available data) and the principle of falsification. In contrast, all of the above principles are satisfied by *a posteriori* methods (Dowling *et al*., 2003). Another advantage of *a posteriori* methods is that in the study of host-parasite associations, parasites are regarded as independent variables whereas *a priori* methods consider parasites to be characters of the host. That makes the results of *a posteriori* methods biologically more realistic. The ontological difference in perspective represented by *a priori* and *a posteriori* approaches reflect changing opinions about the nature of parasitism and the evolutionary independence of hosts and parasites (Dowling *et al*., 2003).

In the opinion of Yang (2009), the most logical approach are event-based methods, provided that they include all scenarios. However, an excess of outcomes from a given analysis can lead to difficulty in finding the actual relationship. Furthermore, both methods do not consider potential phylogenic sampling and reconstruction error. Instead, global-fit statistical methods account for such error and provide a good overview of congruence in a

relationship. Where they fall short, however, is in linking with specific coevolutionary scenarios. Still all types of methods can be improved, and combining the advantages of existing techniques into a single approach will substantially help in future cophylogenetic analyses of all phylogenetic relationships (Yang, 2009).

Molecular-based phylogenetic studies of hosts and parasites are increasingly common in the literature. Phylogenetic comparisons of hosts and their symbiotic associates offer the potential for studies that extend well beyond simple documentation of cophylogeny. If data gathered about the host and its associate (and trees resulting from those data) are statistically independent, show significant cophylogeny, and are based on homologous molecular markers. Then timing of cladogenetic events and possible differences in rate of molecular evolution in the hosts and associates can be estimated. Such studies have the potential to elucidate broad evolutionary processes that influence rates of molecular evolution across distantly related taxa (Hafner *et al*., 1994; Light and Hafner, 2007).

For phylogenetic reconstruction at the species level, the use of several independent gene trees is recommended to overcome the effect of stochastic sorting of ancestral polymorphisms. In addition, a comparison of nuclear and mitochondrial (mt) DNA genealogies can be a powerful tool for detecting hybridization (Moore, 1995). Molecular phylogenetic analysis based on multiple mitochondrial and nuclear loci would be especially useful when obvious incongruence is found between phylogenetic hypotheses from single molecular (*e.g*., mtDNA) markers and non-molecular (*e.g*., morphological) evidence at an early stage of a phylogenetic study (Normark and Lanteri, 1998; Sota and Vogler, 2001). It is believed that on the basis of only one marker, we can not reliably reconstruct the phylogeny of the species. Some studies revealed that mitochondrial DNA (mtDNA) has been found to be a 'powerful subject for evolutionary studies' (Dowton and Austin, 1995; Huelsenbeck *et al*., 1997; Dixon, 2002). This is due to the fact that the mitochondrial DNA is more susceptible to introgression than the nuclear DNA (Coyne and Orr, 2004). However, the other studies proved that mtDNA haplotypes do not coincide with the morphological species. The phylogeny of the three nuclear markers, at least when combined in a simultaneous phylogenetic analysis, is more congruent with the morphologically recognized species than is the mtDNA. Importantly for the test of the type-switching hypothesis, the various nuclear genes also exhibit incongruence with one another, possibly suggesting an overall high amount of incongruence between the various parts of the genome (Su *et al.,* 1996; Sota and Vogler, 2001).

Page and colleagues (1996) suggested several basic pre-requisites for the study of host-parasite associations, among which they singled out: an adequate alpha-taxonomy of both host and parasites, accurate phylogenies of host and parasites, and exhaustive sampling of clades or molecular phylogenies based on comparable genes (Page *et al*., 1996). Unfortunately, accurate phylogenies of both hosts and parasites are not always available, but some of the methods described above (*e.g.,* maximum likelihood methods) are able to generate robust data despite phylogenetic inaccuracies (Huelsenbeck *et al*., 2000; Paterson and Banks, 2001). The disadvantage of maximum likelihood methods, though, is that they are limited to the study of molecular data and cannot take into consideration any other kind of data. Nevertheless, molecular divergence itself is very useful in testing scenarios of phylogenetic events. It not only provides an opportunity to examine comparative rates of molecular evolution in host and parasites, but it also makes it possible to gain knowledge of the timing of different evolutionary events (Paterson *et al*., 2000).

As we have shown, the study of host-parasite interactions has benefited from some major advances in analysis. However, none of the existing methods is ideal. In reconstructing the phylogenetic history of a host-parasite association, the above-mentioned methods can only yield possible scenarios of the common history of the studied species. Researchers must keep the shortcomings of different methods in mind when interpreting the results and especially when comparing results and interpretations obtained by different methods.

References
Ackerly DD. Taxon sampling, correlated evolution, and independent contrasts. Evolution 54: 1480-1492, 2000.

Agrawal AF, Lively CM. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. Evol. Ecol. Res. 4: 79-90, 2002.

Balbuena AA, Míguez-Lozano R, Blasco-Costa I. PACo: a novel procrustes application to cophylogenetic analysis. PLoS One 8: e61048, 2013.

Banks JC, Paterson AM. Multi-host parasite species in cophylogenetic studies. Int. J. Parasitol. 35: 741-746, 2005.

Baum BR, Johnson DA. The application of cophylogenic tools to gene sets. Taxon 59: 1843-1852, 2010.

Behnke JM, Barnard CJ, Wakelin D. Understanding chronic nematode infections: evolutionary considerations, current hypotheses and the way forward. Int. J. Parasitol. 22: 861-907, 1992.

Brooks DR. Hennig's parasitological method: a proposed solution. Syst. Zool. 30: 229-249, 1981.

Brooks DR. Historical ecology: a new approach to studying the evolution of ecological associations. Ann. Missouri Bot. Gard. 72: 660-680, 1985.

Brooks D. Analysis of host-parasite coevolution. Int. J. Parasitol. 17: 291-297, 1987.

Brooks D. Macroevolutionary comparisons of host and parasite phylogenies. Ann. Rev. Ecol. Syst. 19: 235-259, 1988.

Brooks DR, McLennan DA. Parascript. Parasites and the language of evolution. Smithsonian Institution Press, Washington and London DC, pp 1-429, 1993.

Charleston MA. Jungles: a new solution to the host/parasite phylogeny reconciliation problem. Math. Biosci. 149: 191-223, 1998.

Conow C, Fielder D, Ovadia Y, Libeskind-Hadas R. Jane: a new tool for the cophylogeny reconstruction problem. Algorithm. Mol. Biol. 5: 16, 2010.

Cooper MD, Alder MN. The evolution of adaptive immune systems. Cell 124: 815-822, 2006.

Cornell S, Desdevises Y, Rigby M. Evolutionary biology of host-parasite relationships: reality meets models. Trends Ecol. Evol. 14: 423-425, 1999.

Coyne JA, Orr HA. Speciation. Sinauer Associates, Sunderland, MA, USA, pp 1-545, 2004.

Cruaud A, Cook J, Da-Rong Y, Genson G, Jabbour-Zahab R, Kjellberg F, et al. Fig-fig wasp mutualism: the fall of the strict cospeciation paradigm? In: Patiny S (ed.), Evolution of plant-pollinator relationships, Cambridge University Press, Cambridge, UK, pp 68-102, 2012.

Hoyal Cuthill J, Charleston M. Phylogenetic Codivergence Supports Coevolution of Mimetic Heliconius Butterflies. PLoS ONE 7(5): e36464. doi:10.1371/journal.pone.0036464, 2012.

Dawkins R, Krebs JR. Arms races between and within species. Proc. Roy. Soc. Lond. B Biol. Sci. 205: 489-511, 1979.

de Vienne DM, Giraud T, Shykoff JA. When can host shifts produce congruent host and parasite phylogenies? A simulation approach. J. Evol. Biol. 20: 1428-1438, 2007.

de Vienne DM, Refrégier G, López-Villavicencio M, Tellier A, Hood ME, Giraud T. Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. New Phytol. 198: 347-385, 2013.

Decaestecker E, Gaba S, Raeymaekers JAM, Stoks R, Van Kerckhoven L, Ebert D, et al. Host-parasite 'Red Queen' dynamics archieved in pond sediment. Nature 450: 870-874, 2007.

Desdevises Y. Cophylogeny: insights from fish-parasite systems. Parassitologia 49: 125-128, 2007.

Desdevises Y, Morand S, Jousson O, Legendre P. Coevolution between Lamellodiscus (Monogenea: Diplectanidae) and Sparidae (Teleostei): the study of a complex host-parasite system. Evolution 56: 2459-2471, 2002.

Dixon M. Cospeciation of albatrosses and their chewing lice: COI and 12S rRNA sequence data reveal different coevolutionary histories in four genera of albatross lice and their hosts. MRes project, University of Glasgow, USA, 2002 (http://taxonomy.zoology.gla.ac.uk/publications/undergraduate/megan.pdf).

Dowling APG, Veller MGP, Van Hoberg EP, Brooks DR. A priori and a posteriori methods in comparative evolutionary studies of host - parasite associations. Cladistics 19: 240-253, 2003.

Dowton M, Austin AD. Increased genetic diversity in mitochondrial genes is correlated with the evolution of parasitism in the Hymenoptera. J. Mol. Evol. 41: 958-965, 1995.

Dovney JC. Host-plant relations as data for butterfly classification. Syst. Zool. 11: 150-159, 1962.

Dunlap PV, Gould AL, Wittenrich ML, Nakamura M. Symbiosis initiation in the bacterially luminous sea urchin cardinalfish Siphamia versicolor. J. Fish Biol. 81: 1340-1356, 2012.

Erler S, Popp M, Wolf S, Lattorff HMG. Sex, horizontal transmission, and multiple hosts prevent local adaptation of Crithidia bombi, a parasite of bumblebees (Bombus spp.). Ecol. Evol. 2: 930-940, 2012.

Felsenstein J. Phylogenies and the comparative method. Amer. Nat. 125: 1, 1985.

Forde SE, Thompson JN, Holt RD, Bohannan BJM. Coeolution drives temporal changes in fitness and diversity across environments in a bacteria - bacteriophage interaction. Evolution 62: 1830-1839, 2008.

Futuyma DJ, Slatkin M. Coevolution, Sinauer Associates, Sunderland, Massachusetts, USA, 1983.

Gandon S, Angnew P, Michalakis Y. Coevolution between parasite virulence and host life-history traits. Amer. Nat. 160: 374-388, 2002.

Hafner M, Sudman P, Villablanca FX, Spradling TA, Demastes JW, Nadler SA. Disparate rates of molecular evolution in cospeciating hosts and parasites. Science 265: 1087-1090, 1994.

Herrera CS, Hirooka Y, Chaverri P. Pseudocospeciation of the mycoparasite Cosmospora with their fungal hosts. Ecol. Evol. 2016 (doi: 10.1002/ece3.1967).

Hird M. Coevolution, Symbiosis and Sociology. Ecol. Econ. 69: 737-742, 2010.

Hoberg EP, Brooks DR, Siegel-Causey D. Host-parasite co-speciation: history, principles, and prospects. In: Clayton DH, Moore J. (eds), Host-parasite evolution: General principles and avian models, pp. 212-235, Oxford University Press, Oxford, UK, 1997.

Hommola K, Smith J, Qiu Y, Gilks W. A permutation test of host-parasite cospeciation. Mol. Biol. Evol. 26: 1457-1468, 2009.

Hougen-Eitzman D, Rausher MD. Interactions between herbivorous insects and plant-insect coevolution. Amer. Nat. 143: 677-697, 1994.

Huelsenbeck JP, Rannala B, Yang Z. Statistical tests of host-parasite cospeciation. Evolution 51: 410-419, 1997.

Huelsenbeck J, Rannala B, Larget B. A Bayesian framework for the analysis of cospeciation. Evolution 54: 352-364, 2000.

Hypsa V. Parasite histories and novel phylogenetic tools: Alternative approaches to inferring parasite evolution from molecular markers. Int. J. Parasitol. 36: 141-155, 2006.

Imber MJ. Origins, phylogeny and taxonomy of the gadfly petrels *Pterodroma* spp. Ibis 127: 197-229, 1985.

Johnson KP, Adams RJ, Page RDM, Clayton DH. When do parasites fail to speciate in response to host speciation? Syst. Biol. 52: 37-47, 2003.

Joron M, Mallet JL. Diversity in mimicry: paradox or paradigm? Trends Ecol. Evol. 13: 461-466, 1998.

Jousselin E, Van Noort S, Berry V, Rasplus JY, Rønsted N, Erasmus JC, Greeff JM. One fig to bind them all: Host conservatism in a fig wasp community unraveled by cospeciation analyses among pollinating and nonpollinating fig wasps. Evolution 62: 1777-1797, 2008.

Kupczok A, von Haeseler A. Comment on "A congruence index for testing topological similarity between trees. Bioinformatics 25: 147-149, 2009.

Lambrechts L, Fellous S, Koella JC. Coevolutionary interactions between host and parasite genotypes. Trends Parasitol. 22: 12-16, 2006.

Legendre P, Desdevises Y, Bazin E. A statistical test for host-parasite coevolution. Syst. Biol. 51: 217-234, 2002.

Light JE, Hafner MS. Cophylogeny and disparate rates of evolution in sympatric lineages of chewing lice on pocket gophers. Mol. Phylogenet. Evol. 45: 997-1013, 2007.

Meier-Kolthoff JP, Auch AF, Huson DH, Göker M. COPYCAT: cophylogenetic analysis tool. Bioinformatics 23: 898-900, 2007.

Merkle D, Middendorf M. Reconstruction of the cophylogenetic history of related phylogenetic trees with divergence timing information. Theor. Biosci. 123: 277-299, 2005.

Moore WS. Inferring phylogenies from mtDNA variation: Mitochondrial-gene trees versus nuclear-gene trees. Evolution 49: 718-726, 1995.

Moreno-Gámez S, Stephan W, Tellier A. Effect of disease prevalence and spatial heterogeneity on polymorphism maintenance in host-parasite interactions. Plant Pathol. 62: 133-141, 2013.

Morrone JJ, Crisci JV. Historical biogeography: Introduction to methods. Annu. Rev. Ecol. Syst. 26: 373-401, 1995.

Nelson LA, Davies KA, Scheffer SJ, Taylor GS, Purcell MF, Giblin-Davis RM, *et al*. An emerging example of tritrophic coevolution between flies (Diptera: Fergusoninidae) and nematodes (Nematoda: Neotylenchidae) on *Myrthaceae* host plants. Biol. J. Linn. Soc. 111: 699-718, 2014.

Noda S, Kitade O, Inoue T, Kawai M, Kanuka M, Hiroshima K, *et al*. Cospeciation in the triplex symbiosis of termite gut protists (*Pseudotrichonympha* spp.), their hosts, and their bacterial endosymbionts. Mol. Ecol. 16: 1257-1266, 2007.

Normark BB, Lanteri AA. Incongruence between morphological and mitochondrial- DNA characters suggests hybrid origin of parthenogenetic weevil lineages (genus *Aramigus*). Syst. Biol. 47: 475-494, 1998.

Norton DA, De Lange PJ. Host specificity in parasitic mistletoes (Loranthaceae) in New Zealand. Funct. Ecol. 13: 552-559, 1999.

Page RDM. Component 1.5. Program and user's manual, University of Auckland, Auckland, New Zealand, 1990.

Page RDM. Maps between trees and cladistic analysis of historical associations among genes, organisms and areas. Syst. Biol. 43: 58-77, 1994.

Page RD, Charleston MA. Trees within trees: phylogeny and historical associations. Trends Ecol. Evol. 13: 356-359, 1998.

Page RD, Clayton DH, Paterson AM. Lice and cospeciation: a response to Barker. Int. J. Parasitol. 26: 213-218, 1996.

Page RDM. Introduction. In: Page RDM (ed.), Tangled trees: Phylogeny, cospeciation, and coevolution. The University of Chicago Press, Chicago, USA, pp 1-21, 2003.

Paracer S, Ahmadjian V. Symbiosis: An introduction to biological associations. Oxford University Press, New York, USA, pp 1-304, 2000.

Paterson AM, Banks J. Analytical approaches to measuring cospeciation of host and parasites: through a glass, darkly. Int. J. Parasitol. 31: 1012-1022, 2001.

Paterson AM, Gray RD. Host-parasite cospeciation, host-switching and missing the boat. In: Clayton DH, Moore J. (eds.), Host-parasite evolution: General principles and avian models, pp. 236-250. Oxford University Press, Oxford, UK, 1997.

Paterson AM, Wallis GP, Wallis LJ, Gray RD. Seabird and louse coevolution: complex histories revealed by 12S rRNA sequences and reconciliation analyses. Syst. Biol. 49: 383-399, 2000.

Read A. The evolution of virulence. Trends Microbiol. 2*:* 73-76, 1994.

Refrégier G, Le Gac M, Jabbour F, Widmer A, Hood M, Yockteng R, *et al.* Cophylogeny of the anther smut fungi and their caryophyllaceous hosts: prevalence of host shifts and importance of delimiting parasite species. BMC Evol. Biol. 8: 100, 2008.

Ricklefs RE, Fallon SM. Diversification and host switching in avian malaria parasites. Proc. Roy. Soc. Lond. B Biol. Sci. 269: 885-892, 2002.

Ronquist F. Reconstructing the history of host-parasite associations using generalised parsimony. Cladistics 11: 73-89, 1995.

Ronquist F. Phylogenetic approaches in coevolution and biogeography. Zoologica Scripta 26: 313-322, 1997.

Ronquist F, Huelsenbeck JP. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572-1574, 2003.

Sasaki A. Host-parasite coevolution in a multilocus gene-for-gene system. Proc. Roy. Soc. Lond. B Biol. Sci. 267: 2183-2188, 2000.

Schardl CLS, Raven KDC, Peakman SS, Tromberg AS, Indstrom AL, Oshida RY. A novel test for

host-symbiont codivergence indicates ancient origin of fungal endophytes in grasses. Syst. Biol. 57: 483-498, 2008.

Shaw MR. Parasitoid host ranges. In: Hawkins BA, Sheehan W (eds), Parasitoid community ecology, pp. 111-144. Oxford University Press, New York, UK, 1994.

Simmons MP. Misleading results of likelihood-based phylogenetic analyses in the presence of missing data. Cladistics 28: 208-222, 2012.

Smith CI, Godsoe WKW, Tank S, Yoder JB, Pellmyr O. Distinguishing coevolutionfromcovicariance in anobligate pollination mutualism: asynchronous divergence in joshua tree and its pollinators. Evolution 62: 2676-2687, 2008.

Sota T, Vogler AP. Incongruence of mitochondrial and nuclear gene trees in the Carabid beetles *Ohomopterus.* Syst. Biol. 50: 39-59, 2001.

Starr MP. A generalized a scheme for classifying organismic associations. Symp. Soc. Exp. Biol. 29: 1-20, 1975.

Stevens J. Computational aspects of host-parasite phylogenies. Brief. Bioinform. 5: 339-349, 2004.

Su Z-H, Tominaga O, Ohama T, Kajiwara E, Ishikawa R, Okada TS, *et al*. Parallel evolution in radiation of *Ohomopterus* ground beetles inferred from mitochondrial ND5 gene sequences. J. Mol. Evol. 43: 662-671, 1996.

Susoy V, Herrmann M. Preferential host switching and codivergence shaped radiation of bark beetle symbionts, nematodes of *Micoletzkya* (Nematoda: Diplogastridae). J. Evol. Biol. 27: 889-898, 2014.

Tellier A, Brown JKM. Stability of genetic polymorphism in host-parasite interactions. Proc. Roy. Soc. Lond. B Biol. Sci. 274: 809-817, 2007.

Thompson JN. The coevolutionary process, University of Chicago Press, Chicago, USA, pp 1-290, 1994.

Van Valen L. A new evolutionary law. Evol. Theor. 1: 1-30, 1973.

Van Veller M, Kornet D, Zandee M. Methodologies for testing hypotheses of causal processes in vicariance biogeography. Cladistics 18: 207-217, 2002.

Woolhouse MEJ, Webster JP, Domingo E, Charlesworth B, Levin BR. Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nat. Genet. 32: 569-577, 2002.

Yang A. Analysis of methodology in cophylogenetic studies. Last published at http://biochem218.stanford.edu/Projects%2020 09/Yang2009.pdf, 2009.